

Distribución de frecuencias de una variable cuantitativa

La información recolectada mediante un instrumento de recolección de datos, es sometido a un análisis estadístico teniendo en cuenta la cantidad de datos y la cantidad de diferentes variables o podemos tener en cada caso, dando lugar a la construcción de cuadro de distribución de frecuencias.

Existen varios tipos distribución de frecuencias.

Distribución tipo I.- En este caso los datos deben cumplir las siguientes condiciones:

- ✓ Cuando se tienen muchas observaciones o datos.
- ✓ Se tienen pocos valores diferentes de la variable, es decir se tiene poca variedad de valores.
- ✓ Esta distribución se utiliza para presentar variables discretas por lo general, sin embargo, también se puede usar para variables continuas.

Remarca: Una característica de construir los cuadros de distribución de frecuencias es resumir con la menor cantidad de filas o categorías la información brindada por los datos.

Para la construcción del cuadro correspondiente a este caso, dispondremos de las siguientes normas:

- ✚ La primera columna se escriben los valores de la variable.
- ✚ En la segunda columna se anota el número de veces que aparece cada valor de la variable. Este número se denomina frecuencia absoluta (n_i).
- ✚ En la tercera columna se anotan las proporciones de aparición de cada valor diferente de la variable o lo que es lo mismo la frecuencia relativa (h_i).
- ✚ En la quinta columna se anotan los valores acumulados, en orden por lo general ascendente, de la frecuencia absoluta El conjunto conforma la frecuencia acumulada absoluta (N_i). Y esta dado por:

$$N_i = \sum_{j=1}^i n_j \quad i = \text{subíndices admisibles}, i \in I$$

desarrollando tenemos:

$$N_1 = \sum_{j=1}^1 n_j = n_1$$

$$N_2 = \sum_{j=1}^2 n_j = n_1 + n_2$$

$$N_3 = \sum_{j=1}^3 n_j = n_1 + n_2 + n_3$$

Así sucesivamente

Para su interpretación se entiende como una frecuencia "menor o igual que"

Remarca: $N_i = \text{cantidad acumulada} = \text{frecuencia absoluta acumulada "menor o igual que"}$

- ✚ En la sexta columna se anotan los valores acumulados, en orden por lo general ascendente, de la frecuencia relativa. El conjunto conforma la frecuencia acumulada relativa (H_i). Y esta dado por:

$$H_i \% = \sum_{j=1}^i h_j \%$$

desarrollando tenemos:

$$H_1\% = \sum_{j=1}^1 h_j\% = h_1\%$$

$$H_2\% = \sum_{j=1}^2 h_j\% = h_1\% + h_2\%$$

$$H_3\% = \sum_{j=1}^3 h_j\% = h_1\% + h_2\% + h_3\%$$

Así sucesivamente

Para su interpretación se entiende como una frecuencia “menor o igual que”

Remarca: También existe varios valores similares a “menor o igual que”, esto dependerá de la expresión que lo define

Por ejemplo:

$$N_i^* = \sum_{j=i}^m n_j, \quad m = \text{numero de filas "mayor e igual que"}$$

El modelo de presentación para este tipo de distribuciones se muestra en el cuadro (3.3)

CUADRO (3.3.)

DISTRIBUCIÓN DE FRECUENCIAS DE LA VARIABLE (X)

VARIABLE X	FRECUENCIA ABSOLUTA (n_i)	FRECUENCIA RELATIVA (h_i)	PORCENTAJE ($h_i\%$)	FRECUENCIA ACUMULADA ABSOLUTA (N_i)	FRECUENCIA ACUMULADA RELATIVA porcentual ($H_i\%$)
X_1	n_1	h_1	$h_1 * 100\%$	N_1	$H_1\%$
X_2	n_2	h_2	$h_2 * 100\%$	N_2	$H_2\%$
X_3	n_3		$h_3 * 100\%$	N_3	$H_3\%$
.....
X_m	n_m		h_m	$N_m = n$	$H_m = 100$
TOTAL	n		100		

Fuente:

Remarca: La construcción de los cuadros se realizará de forma ascendente

Ejemplo:.. datos que representan números (cuantitativos)

El Departamento de Biología ha realizado un estudio acerca del peso (Kg), un grupo de estudiantes de un colegio de la ciudad. Para mayor facilidad se ha decidido tomar los valores sin la parte decimal, los resultados se muestran a continuación

58	55	63	58	57	67	63
67	58	67	67	57	58	67
67	58	67	55	58	63	63
58	63	55	58	58	57	58
57	63	63	58	58	63	55

a) Tabule los datos

Cuadro # 3
Tabla de distribución de frecuencias
Peso de estudiantes (X)

X_i	n_i	h_i	$h_i \%$	N_i	$H_i\%$
55	4	0.1143	11.43	4	11.43
57	4	0.1143	11.43	8	22.86
58	12	0.3428	34.28	20	57.14
63	8	0.2286	22.86	28	80
67	7	0.2	20	35	100
Total	35	1	100		

Interpretación:

$N_3 = 20$, la cantidad de personas con un peso con un peso menor o igual que 58 kg

$H_4\% = 80\%$, el porcentaje de personas con un peso menor o igual que 63 kg

Preguntas

¿Qué proporción de estudiantes tiene un peso de 67 kg? Res: 0.2

¿Qué porcentaje de estudiantes tienen un peso de por lo menos 58 Kg? 77.15%

¿Qué cantidad de estudiantes tienen un peso menor que 63 kg? Res: 20

¿Qué porcentaje de estudiantes tiene un peso de a lo mucho 63 Kg? 80%

¿Cómo realizamos la tabla?

Primero, aplicamos esta tabla porque no existen muchos datos y cada dato se repite varias veces

En la primera columna “PESOS” (X), y manejamos el siguiente criterio: escribimos las distintas categorías, es decir, los distintos pesos.

En la segunda columna, es denotado por n_i , que significa la cantidad de veces que se repite cada categoría, por tanto tenemos los siguientes resultados

Ejemplo. - Una empresa que produce tarjetas ha reportado que las ventas del mes anterior de sus 36 empleados es la siguiente (expresado en miles de unidades):

1	5	8	3	5	1	8	3	3	6	6	8
5	1	3	8	1	5	5	8	3	3	5	5
8	1	8	3	6	5	1	3	8	1	5	1

Considere como una columna adicional en el cuadro dada por la expresión:

$$N_i^* = \sum_{j=i}^m n_j, \quad m = \text{numero de filas, dado por interpretacion "mayor e igual que"}$$

a) Construya una tabla de distribución de frecuencias.

b) Que proporción de empleados no supero las 3000 tarjetas vendidas Res: 0.4444

c) La empresa otorgara un bono a los empleados que hayan vendido al menos 6000 tarjetas ¿Cuántos empleados recibirán el bono? 11

X_i	n_i	h_i	$h_i \%$	N_i	$H_i\%$	N_i^*
1	8	0.2222	22.22%	8	22.22%	36
3	8	0.2222	22.22%	16	44.44%	28
5	9	0.25	25%	25	69.44%	20
6	3	0.0833	8.33%	28	77.77%	11
8	8	0.2222	22.22%	36	100%	8
Total	36	1	100%			

Distribución de frecuencias tipo II. Este tipo de tratamiento estadístico se utiliza preferentemente para variables continuas o también cuando se han efectuado muchos valores y son de valores distintos.

Para la construcción de frecuencias de tipo II se sigue el siguiente proceso:

- ✓ Paso 1. Se determina el numero de estratos o clases o filas, el cual puede ser fijado arbitrariamente (según las necesidades del estudio o de investigación), mediante la expresión

$$\text{numero de filas} = m \approx \sqrt{n} \text{ (redondeado) el resultado debe ser entero}$$

donde n = cantidad de observaciones

m = numero de estratos (numero de filas del cuadro)

- ✓ Paso 2. Se calcula el recorrido o rango de la variable, la cual está dada por

$$r = \text{recorrido} = x_{\max} - x_{\min}$$

donde: x_{\max} = valor máximo, x_{\min} = valor mínimo.

- ✓ Paso 3. Se determina la longitud del intervalo de clase (de cada fila). Esta distancia puede ser constante o no, en el caso de tener intervalos de longitud constante, esta se calcula dividiendo el valor del recorrido por el valor del número de estratos, es decir,

$$\text{longitud de cada fila} = c_i = \frac{r}{m}$$

- ✓ Paso 4. Se aplica el siguiente formato.

INTERVALOS DE CLASE $[L_{i-1} - L_i)$	MARCA DE CLASE X	FRECUENCIA ABSOLUTA (n_i)	Frecuencia Relativa (h_i)	FRECUENCIA RELATIVA porcentual ($h_i\%$)	FRECUENCIA ACUMULADA ABSOLUTA (N_i)	FRECUENCIA ACUMULADA RELATIVA ($H_i\%$)
$[x_{\min} = L_0 \xrightarrow{+c_i} L_1)$	X_1	n_1		h_1	N_1	H_1
$[L_1 \xrightarrow{+c_i} L_2)$	X_2	n_2		h_2	N_2	H_2
$[L_2 \xrightarrow{+c_i} L_3)$	X_3	n_3		h_3	N_3	H_3

$[L_M \xrightarrow{+c_i} L_R = x_{\max}]$	X_m	n_m		h_m	$N_m = n$	$H_m = 100$
TOTAL		n		100		

Remarca:

a) Se denomina L_{i-1} = límite inferior, y a L_i = límite superior.

b) para determinar los intervalos, primero se coloca el mínimo valor como límite inferior del primer intervalo, luego se le suma la longitud de clase y se obtiene el límite superior del mismo intervalo, luego se copia el límite superior, pero ahora como límite inferior del segundo intervalo, luego se repite el proceso.

Nótese que:

$$X_i = \text{Marca de clase} = \frac{L_{i-1} + L_i}{2} \text{ en la fila correspondiente}$$

Remarca:

- Para valores de c_i, L_{i-1}, L_i, X_i , usaremos un dígito mas de presicion que tienen los datos, es decir, si los datos consideran valores con un digito despues del punto decimal, entonces los valores que se mencionó usaran dos dígitos despues del punto decimal.
- El limite superior del ultimo intervalo debe ser igual al máximo valor de los datos, por lo cual directamente se escribira dicho valor em la posicion indicada.

$n_i = \text{Frecuencia absoluta} = \text{cantidad de valores que existe dentro el intervalo a consideración}$

$$h_i = \text{frecuencia relativa} = \frac{n_i}{n}$$

$$h_i\% = \text{porcentaje} = \frac{n_i}{n} * 100\%$$

$$N_i = \sum_{j=1}^i n_j = \text{frecuencia absoluta acumulada "menor o igual que"}$$

$$H_i\% = \sum_{j=1}^i h_j\% = \text{frecuencia relativa acumulada "menor o igual que"}$$

3. Si tenemos los siguientes datos:

1	1.5	1.6	1.6	1.7	1.8	1.9	2	3.1	3.2
3.7	3.7	3.8	4	4.5	5	5	6	7	8
9	10	11	13						

a) Tabular u organizar los datos

Cuadro # 2
Tabla de distribución de frecuencias

$[L_{i-1} - L_i)$	X_i	n_i	h_i	$h_i (\%)$	N_i	$H_i (\%)$
[1 - 3.4)	2.2	10	0.4167	41.67	10	41.67
[3.4 – 5.8)	4.6	7	0.2917	29.17	17	70.84
[5.8 – 8.2)	7	3	0.125	12.5	20	83.34
[8.2 – 10.6)	9.4	2	0.0833	8.33	22	91.67
[10.6 – 13]	11.8	2	0.0833	8.33	24	100
TOTAL		24	1	100		

¿Cómo realizamos la tabla?

Primero, aplicamos esta tabla porque no existen muchos datos, pero casi todos los datos son distintos (si intentáramos la tabla precedente obtendríamos muchas filas lo cual no sería conveniente)

En la primera columna (X), y manejamos el siguiente proceso:

- Determinamos la cantidad de datos $n = 24$ (cantidad de datos)
- Determinamos el numero de filas que tendrá mi tabla: $m = \sqrt{n} = \sqrt{24} = 4.89 \approx 5$ (m, significa el numero de filas, por tanto tiene que ser entero por eso se redondea)
- $X_{\max} = 13$, significa el valor máximo
- $X_{\min} = 1$, significa el valor mínimo
- $recorrido = r = x_{\max} - x_{\min} = 13 - 1 = 12$
- $C = \frac{X_{\max} - X_{\min}}{m} = \frac{r}{m}$, que en nuestro caso obtendríamos $C = \frac{13-1}{5} = \frac{12}{5} = 2.4$
- Empezamos a construir, primero se coloca en la primera fila el valor mínimo, es decir 1, para encontrar el otro extremo de mi intervalo sumo mas "C", es decir, $1 + 2.4 = 3.4$, para la segunda fila copio 3.4 y sumo mas "C", es decir, $3.4 + 2.4 = 5.8$, y así sucesivamente obtengo los extremos de mi intervalo 8.2, 10.6 y 13 (13 es el valor máximo de mis datos).

En la segunda columna, se encuentra sumando los dos extremos de mi intervalo y dividiendo entre dos, es decir, $X_i = \frac{L_{i-1} + L_i}{2}$, para la primera fila tenemos $X_1 = \frac{1+3.4}{2} = 2.2$, y así sucesivamente

En la tercera fila, es denotado por n_i , que significa la cantidad de veces que se repite cada categoría, por tanto tenemos los siguientes resultados

Ejemplo.- El contenido de nicotina, en miligramos para 20 cigarrillos de cierta marca se registraron de la siguiente manera:

1.3	2.55	2.4	1.6	1.55	1.77	2.05	1.43	3.05	2.55
1.87	2.69	1.64	1.45	1.75	1.9	2.48	3.02	3.44	2.75

a) Construya el cuadro de distribución de frecuencias, considerando una nueva columna denominada b) El porcentaje de cigarrillos que tiene hasta 1.7 miligramos de nicotina

Cuadro #
Tabla de distribución de frecuencias

$[L_{i-1} - L_i >$	X_i	n_i	h_i	$h_i (\%)$	N_i	$H_i (\%)$
TOTAL		20	1	100%		